Dynamic Power Optimization

Concept

Faster is not always faster for HPC

HPC Applications use

- Processors
- Memory
- Network
- I/O
- GPUs

Nearly impossible to always balance all of them

At Runtime: An Online Approach

Zero-knowledge applications

- No initial knowledge of applications
- No access (nor modification) of source code
- No assumption on application behavior (regularity for example)

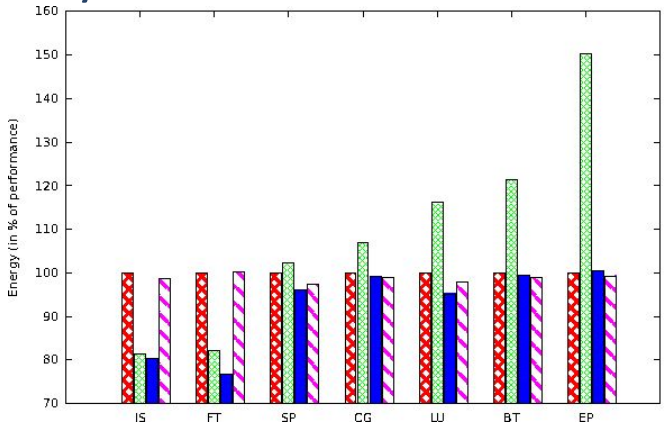
1. *Category* detection at runtime (cpu-bound, memory-bound, …)
2. Category-related leverages applied at runtime (DVFS, ADLR, …)

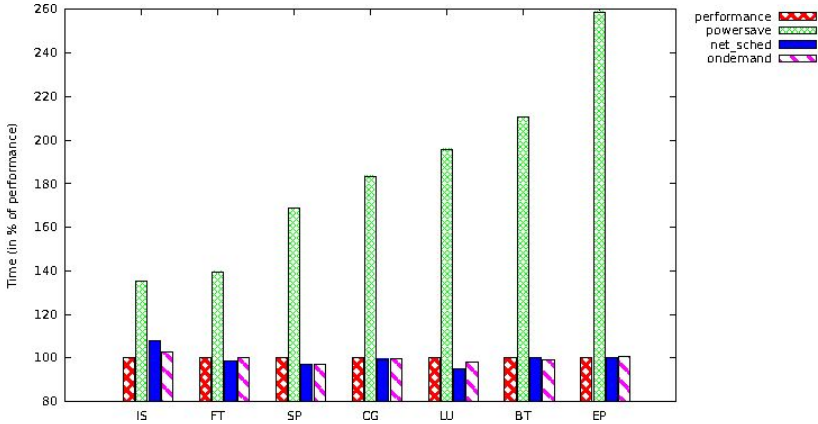
Purely reactive approach

Hypothesis: Only Computation and Communication

- Computation: DVFS at maximum frequency
- Communication: Lower frequency

Decision every 100ms





Currently working on holistic version (inc. network, GPU, …) based on a Decision-Tree fully automatic workflow

Holistic approach

Relax hypothesis on type of applications

1. Monitoring
   a. Hundreds of possible monitored values
   b. PCA to detect the most relevant to the decision
2. Application abstraction
   a. The servers consumes depending on resource consumption not on actual work
   b. Detection of application resource consumption profile
3. Categorization
   a. Detection of the bottleneck (e.g. CPU, Memory, Disk, I/O)
4. Leverages
   a. Reduction of the power consumed for elements unrelated to bottleneck

Monitoring

PCA (Principal Component Analysis), LLCRIR (Last Level Cache References per Instruction Ratio)

- Set of benchmarks with known behavior
  - NAS Parallel Benchmarks
- Analysis of Hardware Performance Counters and system values
  - PCA, LLCRIR, …
- Output:
  - Threshold based labels

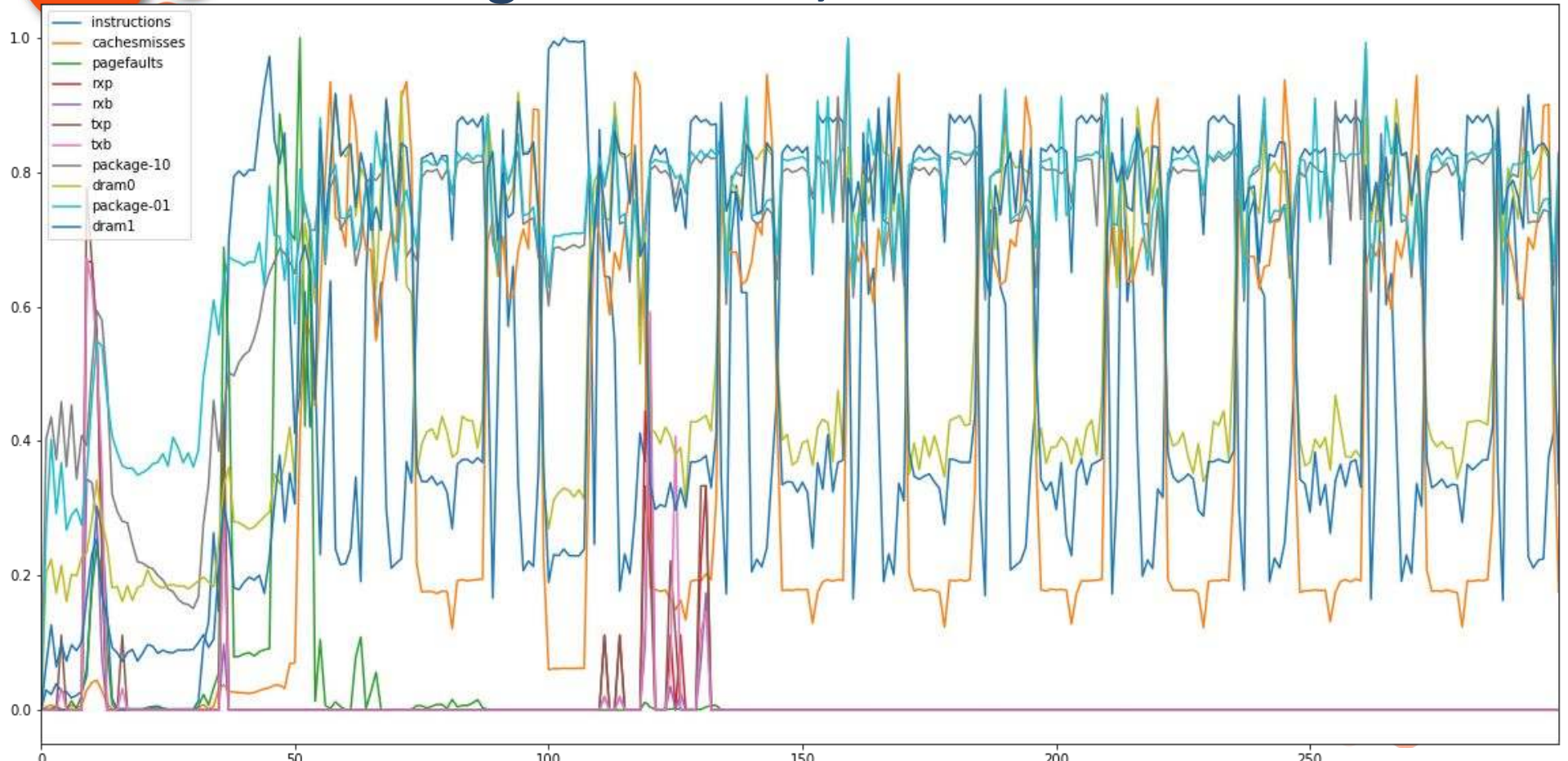
Sensors selected from PCA for phase characterisation	**Associated label**
cache_ref & cache_misses branch_misses or branch_ins	compute-intensive
no IO related sensor	communication / IO intensive
branch_misses & hardware_ins or branch_ins	mixed
hardware_ins & cache_ref or cache_ins	memory-intensive

Online phase detection

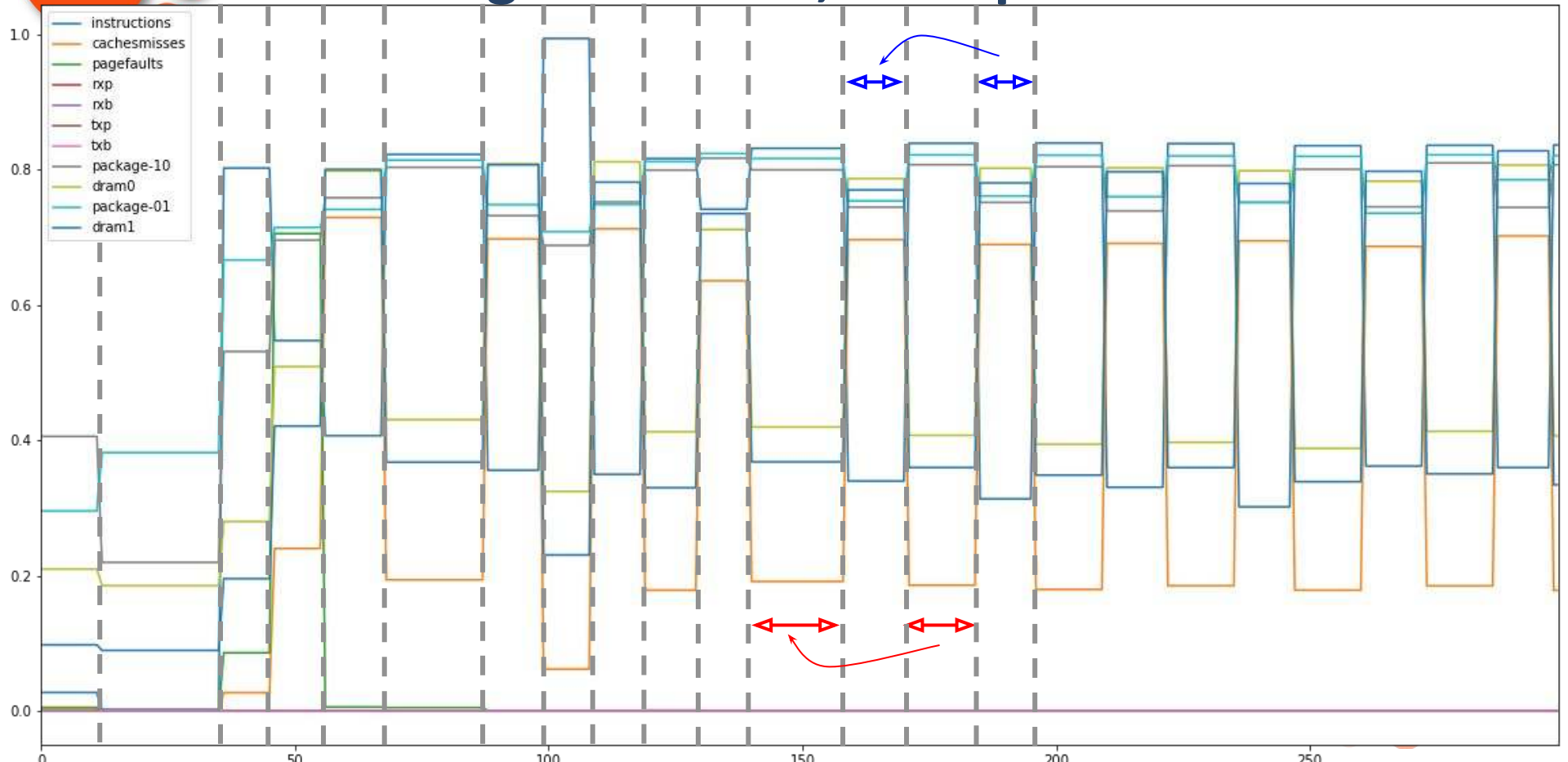
Phase: “Stable resource consumption during a long time (relatively to monitoring frequency)”

- Monitoring of system values, RAPL and hardware performance counters (frequency of 1Hz, 10Hz, depends on the context)
- When values are stable, consider a new phase starts
  - Phase: *[monitored values, duration]*
- Label the new phase (using a *reference vector*)
  - Compare with past phases
  - Check direct label
- Can be used for:
  - Estimation of phase duration
  - Detection/learning of best leverages

LU Diagonalization, raw values



LU Diagonalization, after phase detection



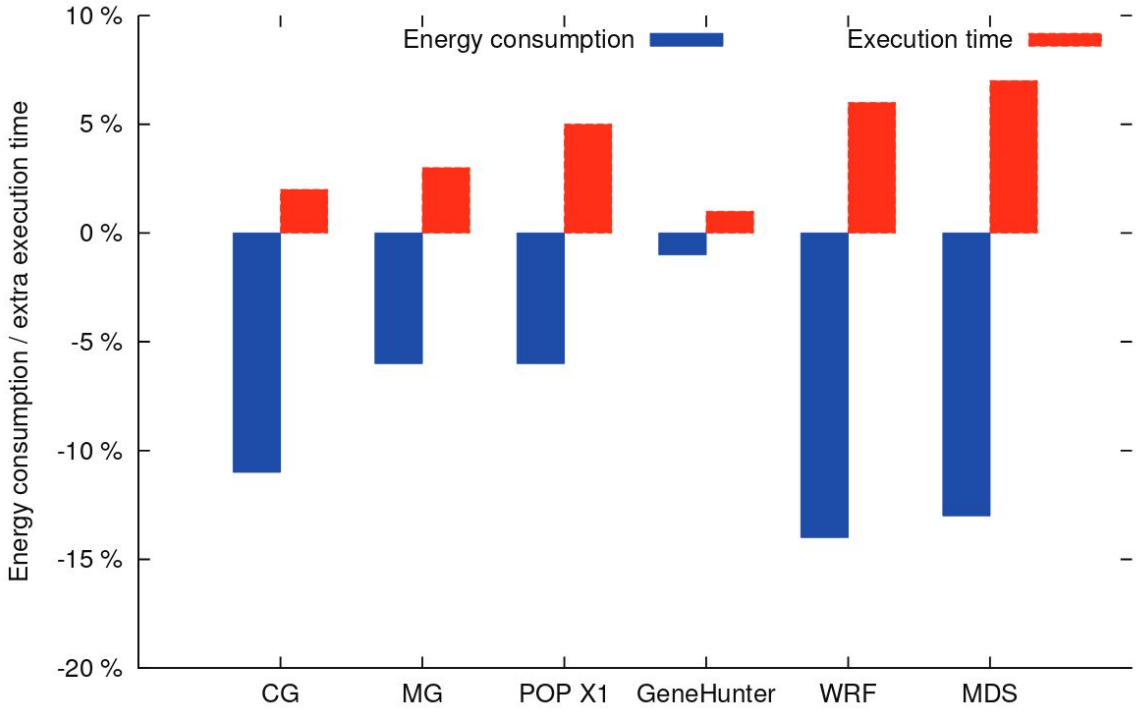
Categorization and Leverages

Phase label	Possible reconfiguration decisions
compute-intensive	switch off unused memory banks; send disks to sleep; scale the processor frequency up; put NICs into LPI mode.
memory-intensive	scale the processor frequency down; decrease disks; or send them to sleep; switch on memory banks.
mixed	switch on memory banks; scale the processor frequency up; send disks to sleep; put NICs into LPI mode.
communication intensive	switch off memory banks; scale the processor frequency down; switch on disks.
IO-intensive	switch on memory banks; scale the processor frequency down; increase disks (if needed).

Results

Energy and performance are antagonistic

- With rule based approach (right)
  - Up to 15% energy savings
- Could be improved with reinforcement learning



References

[1] Zero-knowledge approach

- Da Costa, Georges, and Jean-Marc Pierson. **Dvfs governor for hpc: Higher, faster, greener.** 23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. IEEE, PDP 2015, https://oatao.univ-toulouse.fr/15211/
- PhD CIFRE with Linaro (Daniel Lezcano) started in 2019

[2] Holistic approach

- Tsafack Chetsa, Ghislain Landry and Lefèvre, Laurent and Pierson, Jean-Marc and Stolf, Patricia and Da Costa, Georges **Application-Agnostic Framework for Improving the Energy Efficiency of Multiple HPC Subsystems**. In: 23rd Euromicro International Conference on Parallel, Distributed and network-based Processing, IEEE, PDP 2015, https://oatao.univ-toulouse.fr/15207/
- Ghislain Landry Tsafack Chetsa, Laurent Lefèvre, Jean-Marc Pierson, Patricia Stolf, Georges Da Costa **Exploiting performance counters to predict and improve energy performance of HPC systems**, Future Generation Computer Systems, Elsevier, 2014, https://oatao.univ-toulouse.fr/12657/

Collaboration possibilities

To be discussed…